

⇒[MetaMap Home](#)
⇒[FAQ Home](#)
⇒[Help/Questions](#)

MetaMap's Default Input: Unformatted ASCII English Free Text

MetaMap's default input is **unformatted ASCII English free text**.

Input records are by default separated by a blank line; however see the other sections in [FAQ #2](#) for additional input formats.

If your input text contains non-ASCII characters, e.g., *Behçet's Disease*, *Sjögren's Syndrome*, *Guillan-Barré Syndrome*, *Ménière's Disease*, *Waldenström Hyperglobulinemic Purpura*, *Canastrão*, *Capitão Chico pig*, *Aksaî Black Pied pig*, etc., MetaMap will abort.

A Java program that converts the most common non-ASCII characters to their closest ASCII equivalents is available [here](#).

The last line of all MetaMap input files must end with a newline character!

Because MetaMap by default processes English text only, MetaMap's UMLS-based datafiles that are used to map text to UMLS concepts include only the English strings from the UMLS sources (i.e., `LAT == "ENG"`). Users needing to include sources in languages other than English in order to process non-English text should create their own UMLS-like database using our Data File Builder suite, available [here](#).